The Photogrammetric Record 20(111): 285–302 (September 2005)

3D BUILDING MODELLING WITH DIGITAL MAP, LIDAR DATA AND VIDEO IMAGE SEQUENCES

YONGJUN ZHANG (yongjun_zhang@hotmail.com)

ZUXUN ZHANG (zxzhang@supresoft.com.cn)

JIANQING ZHANG (jqzhang@supresoft.com.cn)
Wuhan University, China

JUN WU (wujun@supresoft.com.cn)
Shandong University of Science and Technology, China

Abstract

Three-dimensional (3D) reconstruction and texture mapping of buildings or other man-made objects are key aspects for 3D city landscapes. An effective coarse-to-fine approach for 3D building model generation and texture mapping based on digital photogrammetric techniques is proposed. Three video image sequences, two oblique views of building walls and one vertical view of building roofs, acquired by a digital video camera mounted on a helicopter, are used as input images. Lidar data and a coarse two-dimensional (2D) digital vector map used for car navigation are also used as information sources. Automatic aerial triangulation (AAT) suitable for a high overlap image sequence is used to give initial values of camera parameters of each image. To obtain accurate image lines, the correspondence between outlines of the building and their line features in the image sequences is determined with a coarse-to-fine strategy. A hybrid point/line bundle adjustment is used to ensure the stability and accuracy of reconstruction. Reconstructed buildings with fine textures superimposed on a digital elevation model (DEM) and ortho-image are realistically visualised. Experimental results show that the proposed approach of 3D city model generation has a promising future in many applications.

KEYWORDS: 3D city models, coarse-to-fine building modelling, hybrid point/line bundle adjustment, texture mapping, video image sequences

INTRODUCTION

THREE-DIMENSIONAL (3D) interactive environments offer intuitive and user-friendly ways to view location-based information, such as 3D city models. A 3D city model is usually composed of descriptions of terrain, streets, buildings, other man-made objects and vegetation. People are becoming increasingly conscious of the applications of 3D city models, such as car navigation and service browsing, tourism and marketing, architecture and town planning, city climate and environmental research. Interest in 3D city model generation and in frequent

Blackwell Publishing Ltd. 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street Malden, MA 02148, USA.

update has greatly increased over recent years (see, for example, Collins et al., 1998; Gruen, 1998; Brenner et al., 2001; Wang and Hanson, 2001).

However, current digital photogrammetric workstations (DPWs) are not the most efficient solution for creating city models, since they are made to measure points from vertical aerial photographs rather than structured objects mainly composed of line segments. Researchers have developed approaches to modelling buildings using a small number of constructive solid geometry (CSG) models (for example, Debevec et al., 1996; van den Heuvel, 2000). Nevertheless, many interactions make the modelling process less efficient.

The main workload of 3D city model generation is still building modelling, although tremendous progress has been made (Gruen and Wang, 1998; Haala and Brenner, 1998; Baillard and Zisserman, 1999; Stilla et al., 2000; Brenner et al., 2001; Gruen et al., 2003). Combining digital elevation model (DEM) or lidar data with existing two-dimensional (2D) vector digital map data for building reconstruction has shown promising results (Brenner, 1999; Stilla and Jurkiewicz, 1999; Früh and Zakhor, 2001; Nakagawa and Shibasaki, 2001; Rottensteiner and Briese, 2003; Sohn and Dowman, 2003; Wu et al., 2003). Brenner (2003) gives a good review of some proposed automatic and semi-automatic building reconstruction systems from laser scanning and images.

Nowadays, 2D vector data for buildings and other man-made objects is usually available in many cities in North America, Europe and Asia. However, car navigation system companies usually cannot use the 2D vector data free of charge; it can be very expensive, especially in Japan. So they produce their own 2D vector data with low geometric precision for car navigation. The development of science and technology is also leading to more demands for 3D visual navigation, which inevitably requires a 3D city model. In order to reduce interactions per building, semi-automatic capture systems are desirable. Moreover, the aim is to have a system reconstructing a considerable number of buildings without any operator intervention at all.

In this paper, several important techniques for 3D building model generation making use of a "coarse-to-fine" strategy are discussed. The general strategy of how to generate 3D city models with low-accuracy 2D vector data, lidar data and video image sequences is presented in the next section. A semi-automatic coarse-to-fine algorithm for building modelling is then addressed. To generate realistic textures of buildings, texture rectification and improvement solutions are next described. This is followed by discussion of the experimental results and performance of the developed system. Discussion and outlook for future developments conclude the paper.

GENERAL STRATEGY

The overall goal is the generation of 3D city models using available 2D vector data with non-negligible geometric errors (usually larger than 3 m in horizontal accuracy), this normally being acceptable for 2D car navigation. Complex superstructures on the roofs of buildings are not of interest in this paper, because they are not important for car navigation.

Although the 2D vector data may have geometric errors of the order of 5 m, it can be used to provide initial values of building outlines for 3D reconstruction. Such data contains no information about building heights, so data obtained from laser scanning (lidar) with 1 m resolution is treated as a necessary source of height information. Usually, building walls are invisible in vertical aerial images. Therefore, three video image sequences (one vertical view of building roofs and two oblique views of building walls) acquired by a helicopter-mounted digital video camera are also used as input data.

First, automatic aerial triangulation (AAT) of the three image sequences has to be carried out to give initial values of camera parameters for each image. Then, an iterative coarse-to-fine process, composed of feature projection, image matching and bundle adjustment, is adopted to reconstruct buildings, as described in more detail later in this paper. Initial building models are obtained from 2D vector data and lidar data. To ensure the stability and accuracy of reconstruction, a hybrid point/line bundle adjustment (van den Heuvel, 2000; Zhang et al., 2004) making use of the image points of ground features and image lines of buildings as observations with collinearity and coplanarity equations, respectively, is adopted for accurate reconstruction. After that, building textures are rectified and improved semi-automatically to generate realistic visualisations.

The aim of this research is to reconstruct 3D city models especially for car navigation applications. To generate DEMs, buildings and other man-made objects have to be removed from the lidar data, so a data filtering technique is used to produce the DEM of the area of interest. For visualisation of the ground, an ortho-image of the region of interest is also essential. To this end, the DEM, camera parameters and the "level" image sequence (namely, the vertically viewed images) are used to generate the ortho-image by digital orthorectification. Reconstructed buildings with fine textures are merged into the textured ground, which is generated from the DEM and the ortho-image. Finally, the generated 3D city model can be browsed at any viewpoint or even walked along the street.

Coarse-to-Fine Building Reconstruction

Automatic Aerial Triangulation

Traditional aerial photographs usually have 60% forward overlap between adjacent photographs within a strip. So there is at most 20% overlap among three images, which means one ground feature point appears at most in three images within a strip. However, image sequences taken with a video camera have a high percentage of overlap (usually higher than 95%) between adjacent images and therefore more image correspondences for each ground feature, and there is only a slight difference between neighbouring conjugate images of the same ground feature. These advantages make the image matching more reliable and more precise than traditional photogrammetry and thus have good potential for 3D reconstruction.

It is known that the vector data available for car navigation is not accurate. Therefore, both camera parameters and building models must be determined in order to obtain fine textures. Automatic aerial triangulation (AAT) is adopted to acquire initial values of camera parameters. Different from traditional analytical aerial triangulation, the image sequence has high overlap and thus a very short baseline between adjacent images. In an image sequence with 90% overlap, one ground feature may appear in 10 images at most. All the image features corresponding to the same ground feature should be matched, so as to enlarge the intersection angle and thus overcome the problems caused by short baselines. Multi-view image matching and forward intersection algorithms have to be developed to overcome these shortcomings.

The three image sequences are processed separately as single strips. A multi-view image matching algorithm is used to determine the conjugate image features. The output of AAT is a free network taking the first image as the reference coordinate system, which is different from the real-world coordinate system. So the camera parameters of the free network need to be transformed into the world coordinate system to give reliable initial values of the image poses. Two images near the beginning and the end of the image sequences are chosen, followed by manual selection of some control points from the 2D vector data overlaid with lidar data. Their corresponding image points are given interactively in the selected images. Thus camera parameters of the two selected images can be obtained by a space resection process. Then all the camera parameters covering the strip can be transformed into the world coordinate system. All image points will be used in the hybrid point/line bundle adjustment to ensure the stability of adjustment.

Initial Geometric Model of Building

The initial geometric model of the building can be determined by 2D vector data and lidar data. Initial values of ground plans of buildings can be extracted from the 2D vector data, while the height of buildings cannot be obtained directly from lidar data due to noise. Furthermore, most building roofs are not flat. When the lidar data is transformed into a pseudo image with height information as the grey value (Fig. 1), it is clear that there are many small complex structures on the rooftop. Considering the aforementioned problems, all shapes of building rooftops are restricted to being flat, corresponding to 2D vector data. To obtain height information for building rooftops, noise and complex shapes have to be removed. An image smoothing technique with Gaussian filter is applied to the pseudo image data. The result is shown in Fig. 2, where the building height can be taken as the average height of the roof outline in the filtered lidar data. The precision of building heights obtained as above is about 1·5 m, which is satisfactory for their use as initial values.

To acquire initial values of building outlines in the image, the low-accuracy building model is back-projected onto the level (vertical) images in accordance with the AAT results. For each building, all images in which the building of interest is fully included are retrieved automatically from the whole image sequence, by comparing the building location with every image range on the ground, which is calculated from image pose parameters. The best image, in which the building of interest is closest to the centre of the image, is auto-selected as well.

After the initial geometric model of the building is projected back onto the relevant images, an initial wire-frame model of the building is displayed in the image, as shown in Fig. 3. As can be seen in the left part of Fig. 3, there are not only offsets but also scale differences between the projected wire-frame and corresponding image lines of the building.

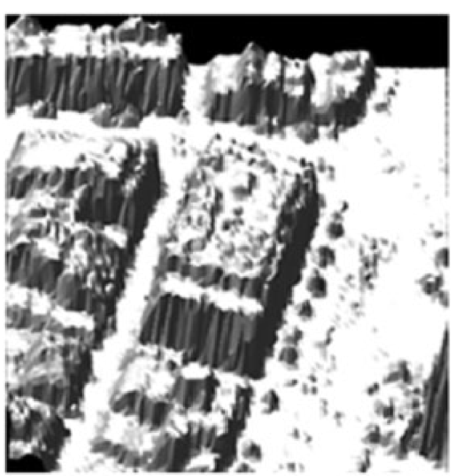

FIG. 1. Lidar data before filtering.

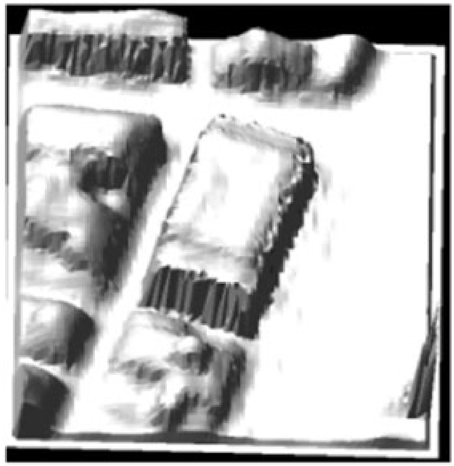

FIG. 2. Lidar data after filtering.

FIG. 3. Initial projection of two buildings.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

The offset and scale factor are partly caused by imprecision of camera parameters because no precise control points were available in AAT. Furthermore, the geometric shape of the building is significantly different from the image because of imprecision of 2D vector data, although offset and scale factor are removed manually, as shown in Fig. 4.

FIG. 4. Imprecision of 2D vector data.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

Coarse-to-Fine Reconstruction by Bundle Adjustment

When the initial building model is ready, image lines corresponding to the projected building outlines will be extracted. The extraction of lines on an image is only performed in the neighbourhood of the projected initial wire-frame of the building. At the beginning, a rectangular image patch with the same length as the projected line and 60 pixels width centred at the projected position of each line is used as target region for possible search areas of image lines. The main purpose is to obtain the strongest line feature to improve the geometric building model. The lines extracted from the image are good enough to improve the geometric model, as shown in Fig. 5. Then, all the lines extracted from the "best" image will be automatically matched from the remaining images by least squares template matching (LSTM) (Gruen, 1985).

However, in some complex situations such as dense buildings and buildings with similar double edges (see rooftop of Fig. 9), the LSTM approach may produce a large number of mismatched lines. To overcome this problem, homologous line matching based on LSTM and trifocal tensors (Hartley and Zisserman, 2000) is applied to obtain more reliable homologous lines in all target images. By the combination of LSTM and trifocal tensors, most corresponding image lines of building outlines can be automatically determined.

After image lines are determined, they can be used for 3D reconstruction. During reconstruction, the flat roof constraint $Z_i - Z_j = \Delta Z_{i,j} = 0$ (Z_i and Z_j are any of the heights of two corners of the same building) is used to restrict the building roofs to being flat. Furthermore, the intersection line of two walls such as L_1 at the upper-right corner of Fig. 5 is always vertical. This information is advantageous to get precise camera orientation angles (ϕ, ω, κ) of each image (Zhang et al., 2003). Constraints of perpendicularity between adjacent lines, for

FIG. 5. Matched image lines of a building.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

example, the line segments P_1P_2 and P_2P_3 of Fig. 5, which are composed of building corners P_1, P_2 and P_3, are also applied in the bundle adjustment. Matched conjugate image points, saved in the process of AAT, are also used to improve the accuracy of camera orientation parameters. Once all the homologous lines for building edges are found across images in one strip, the geometric building model is improved with hybrid point/line bundle adjustment, which is presented by Zhang et al. (2004). The general principle of hybrid point/line bundle adjustment is to use both image points and image lines for 3D reconstruction, with collinearity and coplanarity equations, respectively. This approach is quite effective especially in 3D reconstruction of urban scenes and industrial objects.

More accurate geometric building models will be achieved as they are projected onto oblique images. Using the same methods mentioned above, a number of building lines are extracted and their homologous lines are found. Once more precise image lines are available from line matching, the bundle adjustment will be repeated, and camera parameters of the image sequences are also updated according to the results of bundle adjustment. As a result a large part of the errors inherent in vector data can be considerably corrected, as shown in the left part of Fig. 6. Furthermore, the nearby building is also much closer to its corresponding image feature after bundle adjustment, as shown in the right part of Fig. 6 (compared with the right part of Fig. 3).

If the initial building model is not good enough because of a very large offset (more than 50 pixels, say) between projected and real image lines, the user should intervene by dragging the projected model approximately over its image. Then, the bundle adjustment process will be repeated until the manually dragged building models are well enough fitted to the corresponding images. After several buildings are reconstructed, user intervention quickly

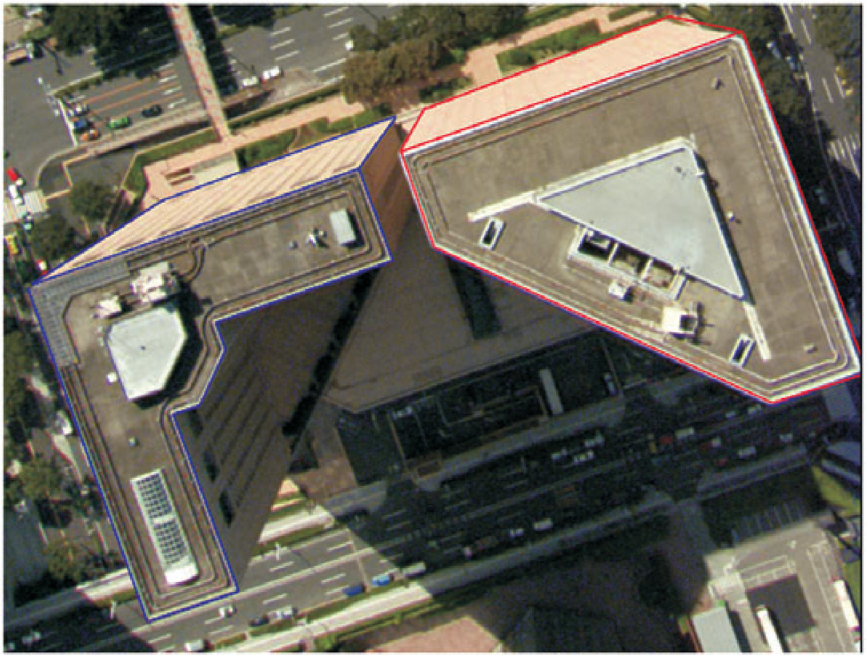

FIG. 6. Intermediate result of coarse-to-fine strategy.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

decreases. Finally, a large number of buildings can be reconstructed automatically and precise camera parameters can be generated.

TEXTURE RECTIFICATION AND IMPROVEMENT

Since one of the most important applications of 3D city models is the generation of realistic visualisations, proper representation of the geometry and texture of buildings is required. When all buildings are successfully reconstructed, the precise camera parameters of each image can also be acquired. With this information, the reconstructed 3D building can be back-projected onto the image and thus textures of building walls can be retrieved. Where a wall appears in multiple images, an image with maximum projected area is finally selected for providing the corresponding texture. Texture for building rooftops is mapped from the level images. It is usually taken from the middle image among all the target images in which it is visible. The geometric resolution of image patches of walls from level images is often lower than that in the oblique ones, so the best texture of the visible wall is auto-selected and mapped from the oblique images with the same strategy as rooftops. Fig. 7 shows a result before and after texture rectification.

Although the aforementioned texture rectification leads to satisfactory results, the method is only applicable to views without occlusions. Unfortunately, there are usually many textures of building walls that are occluded either by other buildings as shown in Fig. 8, or by tall trees in the street. These textures have to be refined to generate realistic 3D city models. To this end, the texture improvement process is done semi-automatically under the control and supervision

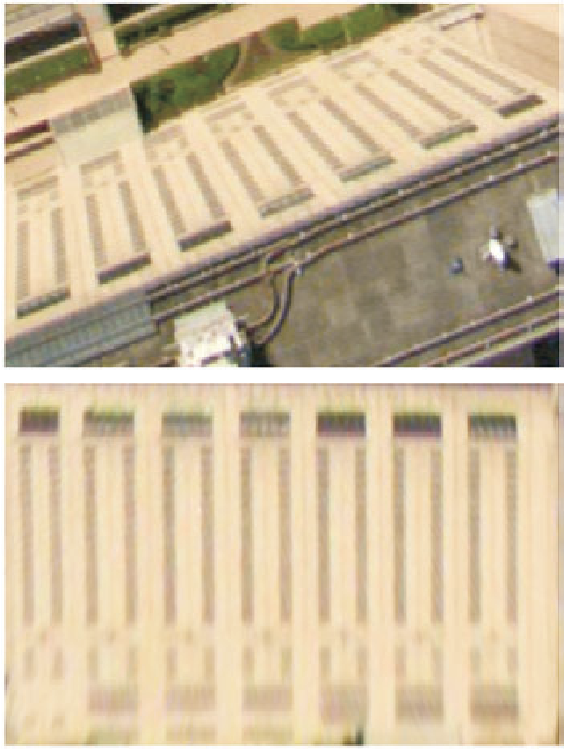

FIG. 7. Texture of a wall before and after rectification.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

FIG. 8. Textures of walls occluded by other buildings.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

of the operator (see Fig. 8). At first, textures are automatically processed by computer. Human interaction is needed only when the visual effects are unsatisfactory. The wall textures to be processed can be selected by the operator without specifying the method. Occlusion restoration and blur removal are applied to textures of the first and third wall. Occlusion restoration is

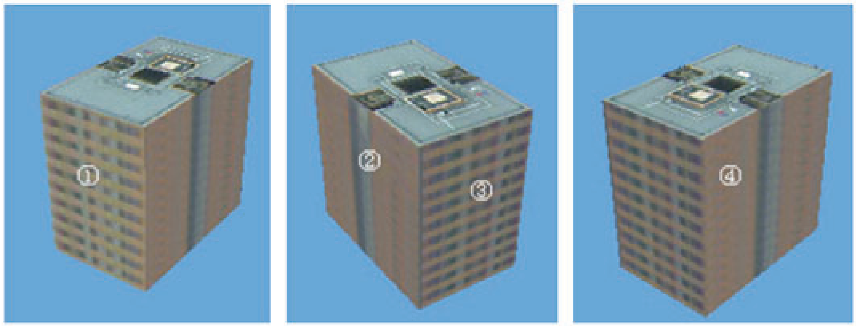

FIG. 9. The improved textures of walls.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

applied to the texture of the fourth wall. The texture of the second wall is duplicated from that of the fourth wall. As can be seen from Fig. 9, the improved textures of walls are much better than the raw ones in Fig. 8.

EXPERIMENTAL RESULTS

A 3D city modelling system has been developed in accordance with the algorithms described above. The essential input data sources are three video image sequences, ground control points, internal camera parameters, lidar data and 2D vector data. The system is composed of several kernel modules, such as AAT (including interior orientation, relative orientation, model linking, absolute orientation and bundle adjustment), initial building model generation, multi-view image matching, coarse-to-fine reconstruction by bundle adjustment, texture mapping and improvement, 3D visualisation and so on. The flow chart of the developed system is shown in Fig. 10. Firstly, the AAT module is called to obtain the external camera parameters of the three image sequences. Initial building models are generated by processing the lidar and 2D vector data simultaneously. Then images of interest for each building are automatically selected followed by line extraction and multi-view image matching. Afterwards, 3D building models are reconstructed by the coarse-to-fine strategy with bundle adjustment. When the best textures of buildings are automatically selected from the image sequences, the texture improvement process is done semi-automatically under the control and supervision of the operator. The 3D models with refined textures can then be visualised realistically.

The developed system has been tested using data for several city blocks taken by a pre-calibrated video camera. All experimental results are very promising. A typical experimental result for a street in Japan, 1·7 km long with about 200 buildings, will be discussed in the following.

Image sequences were taken by a digital video camera with resolution of 1920×1200 pixels mounted on a helicopter. The ground resolution of images is about 0·2 m, which means that buildings smaller than $380\,m \times 240\,m$ can be fully included within a video frame. The camera is fixed in a stabilising mount, which can keep the camera orientation more stable than the helicopter itself while flying. There is almost no vibration of the camera rotation angles. There are in total three image sequences acquired by three flying routes as shown in Fig. 11, two sequences (image sequences 1 and 2 obtained by flying routes 1 and 2, respectively) with oblique view-angles of about 45° to acquire the images of building walls, the other one (image

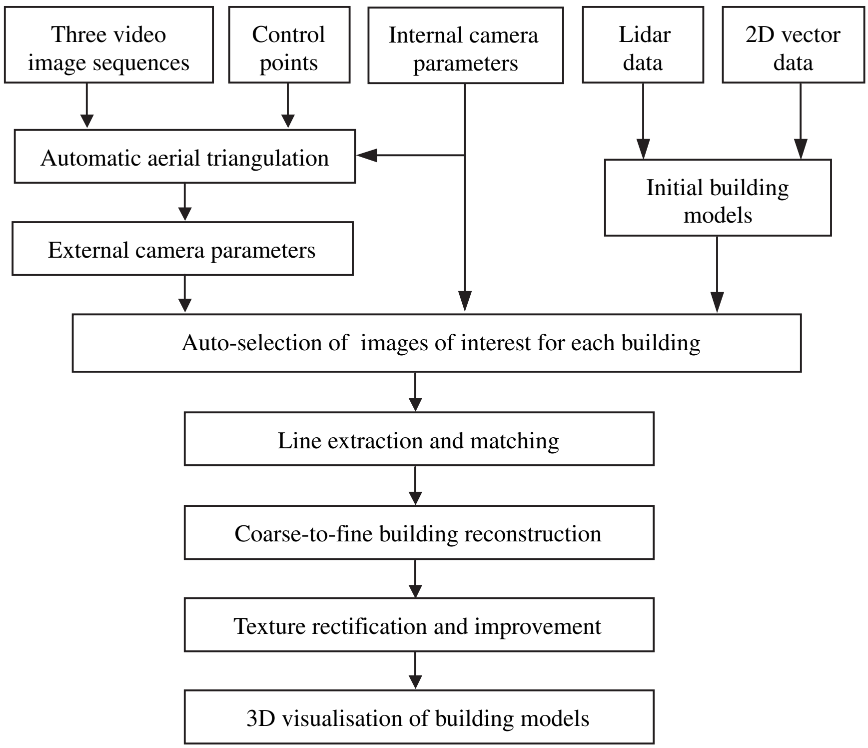

FIG. 10. Outline of 3D building modelling and texture mapping.

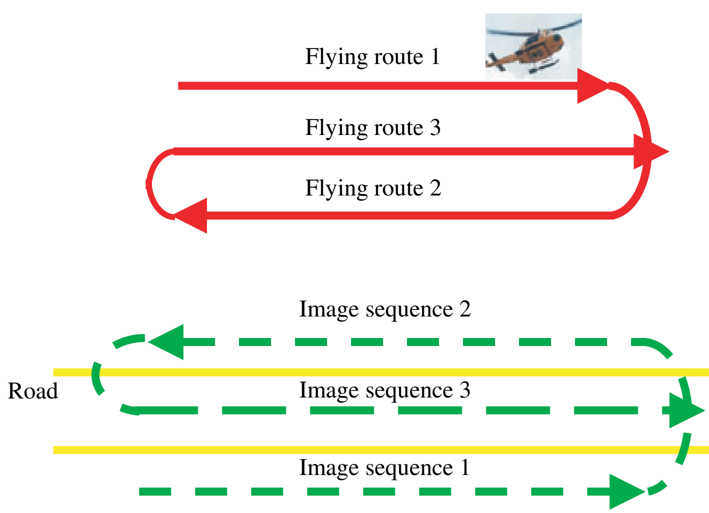

FIG. 11. Three flying routes of the helicopter.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

sequence 3 obtained by flying route 3) with optical axes looking vertically downwards to obtain the images of roofs and roads. So images of the first and second sequences are oblique (called oblique images in the following), while images of the third sequence (vertical view) are

FIG. 12. Part of the first oblique image sequence.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

FIG. 13. Part of the second oblique image sequence.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

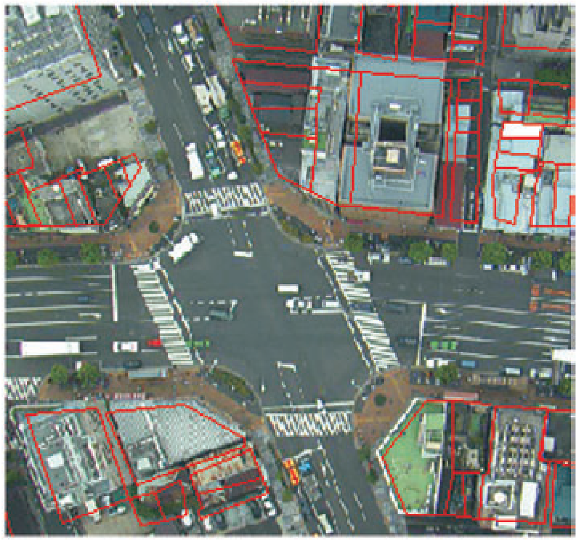

FIG. 14. Part of the level (vertical) image with vector data.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

described as "level". The three image sequences are extracted from the obtained video data with 90% overlap between adjacent images. There are more than 1000 images in total, extracted from the three video sequences. Part of two oblique images and a level image are shown in Figs 12, 13 and 14, respectively.

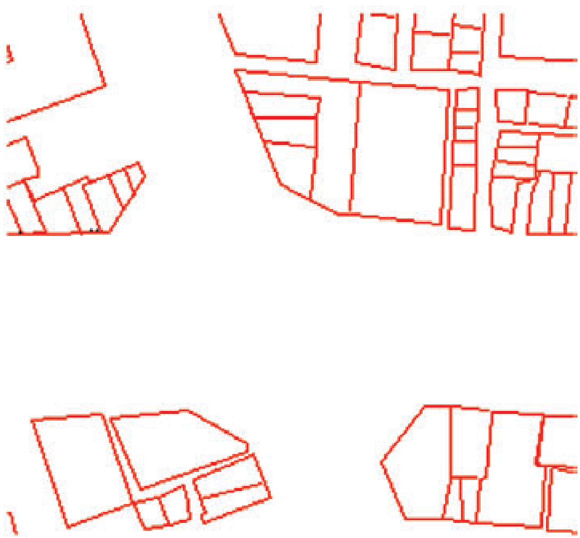

FIG. 15. 2D vector data of a road intersection.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

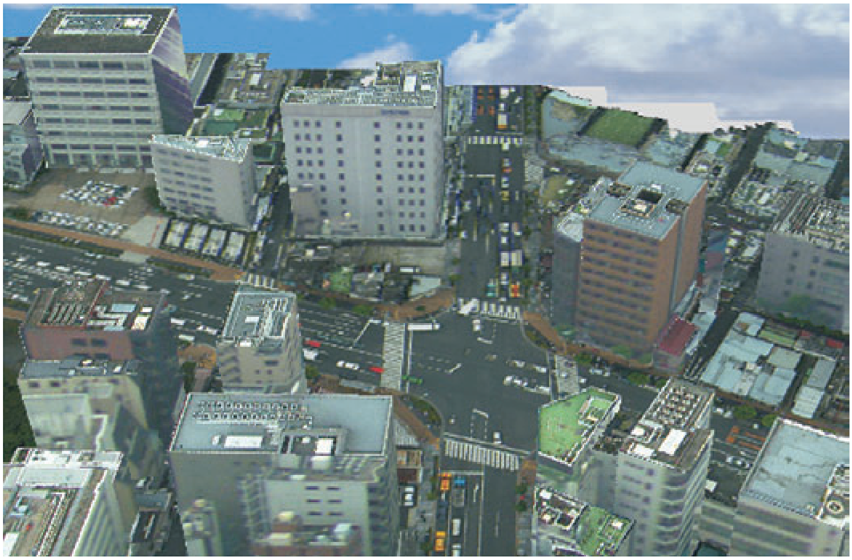

FIG. 16. Flying overview of the reconstructed 3D scene of a road intersection.
This figure appears in colour in the electronic version of the article and in the plate section at the front of the printed journal.

As mentioned in the second section, the lidar data with ground resolution of 1 m is acquired by a helicopter-mounted laser scanner. A similar example showing the quality of the lidar data used is shown in Fig. 1. It is also used to produce a DEM by removing man-made objects and vegetation with data filtering techniques similar to image filtering with a Gaussian kernel.

The available 2D digital vector data, which is only used for car navigation, is obtained by detailed survey. Precision of the 2D vector data is about 5 m, which is definitely not good enough to present precise 3D building models, as can be seen from Fig. 14. Part of the 2D digital vector data of a road intersection is shown in Fig. 15. In the 2D vector data used, some buildings are missing, some building footprints are not enclosed and some are seriously distorted. About 30 buildings are excluded from building reconstruction and only 170 buildings of the total of 200 are considered in this study. These buildings are visible in 660 images.

All the 170 buildings are successfully reconstructed by the developed system. Precision of reconstructed 3D building models is better than 0·15 m, comparable to the image resolution. Textures of walls are also rectified and improved. A realistic flying overview of the generated 3D city model of a road intersection is shown in Fig. 16.

The time spent on AAT and 3D reconstruction of the proposed approach including texture rectification and improvement is listed in Table I. In this table, "automatic" means the time that the computer runs automatically without any human interaction. "Interactive" means the

TABLE I. Time spent on AAT and 3D building reconstruction.

Process	*Number*	*Automatic*	*Interactive*
AAT	660 images	8·5 h	8 h
3D reconstruction	170 buildings	2 h	8 h

TABLE II. Comparison of efficiencies between pure manual human labour and the semi-automatic system.

Testing area	*Reconstruction method*	*Number of buildings*	*Total time spent*	*Efficiency*
A city in China	Human (manual) labour	6000	600 working days	10 buildings/day
A street in Japan	Semi-automatic system	170	17 h	80 buildings/day

time spent by an operator for interactively performing the AAT and 3D reconstruction process. As can be seen, 170 buildings are successfully reconstructed within 8 h after AAT. The average time spent to reconstruct one building is less than 3 min. That means 20 buildings can be reconstructed within 1 h if camera parameters are available. Considering the time spent for AAT, data conversion, DEM and ortho-image generation, it takes 17 h to reconstruct the 3D city model of the test area (that means 10 buildings per hour). It may be concluded that about 80 buildings can be reconstructed per person every working day (8 h).

A project is in progress to generate a 3D model of a city in China manually with human labour. The complexity of buildings is similar to that of the test area in Japan. Ten people spent 600 working days to reconstruct 6000 buildings, which means only 10 buildings were reconstructed every working day per person. Clearly, the efficiency of the automated system is eight times higher, as shown in Table II.

DISCUSSION

In this paper, an effective approach to generate 3D building models and hence 3D city models using a coarse-to-fine strategy is proposed. Three video image sequences, coarse 2D vector data of buildings and lidar data are used as information sources. Efficiency of 3D building reconstruction has been shown to be eight times higher than traditional manual methods fully relying on human labour. If precise 2D vector data is available, the reconstruction process will be much easier and quicker. Experiments in 3D city model generation show that the proposed approach has a promising future in many practical applications.

However, the proposed system still requires further improvement in order to achieve more efficient generation of 3D city models. There are often trees on both sides of roads, which need to be reconstructed for realistic visualisation. In addition, tree occlusions and cars on the roads must be removed, while road markings should be retained. For a true 3D city model, complex superstructures on the building roofs also need to be reconstructed, although there is no information available in the 2D digital vector data. One possible solution for superstructures is semi-automatic reconstruction from lidar data and image sequences. How to combine conventional large format aerial survey image data with the video camera sequences also needs further investigation.

ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (NSFC) under project No. 40301041. The authors would like to thank MapMaster (Japan) for providing the test data. Heartfelt thanks are also given for the comments and contributions of anonymous reviewers and members of the Editorial team.

REFERENCES

BAILLARD, C. and ZISSERMAN, A., 1999. Automatic reconstruction of piecewise planar models from multiple views. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 559–565.

BRENNER, C., 1999. Interactive modelling tools for 3D building reconstruction. *Photogrammetric Week '99* (Eds. D. Fritsch and R. Spiller). Wichmann, Heidelberg. 380 pages: 23–34.

BRENNER, C., 2003. Building reconstruction from laser scanning and images. *Proceedings of the ITC Workshop on Data Quality in Earth Observation Techniques*, Enschede, the Netherlands. 8 pages (on CD-ROM).

BRENNER, C., HAALA, N. and FRITSCH, D., 2001. Towards fully automated 3D city model generation. *Automatic Extraction of Man-Made Objects from Aerial and Space Images III* (Eds. E. P. Baltsavias, A. Gruen and van Gool). A. A. Balkema, Lisse, the Netherlands. 415, 47–57.

COLLINS, R., JAYNES, C., CHENG, Y., WANG, X., STOLLE, F., SCHULTZ, H., RISEMAN, E. and HANSON, A., 1998. The ascender system: automated site modelling from multiple aerial images. *Computer Vision and Image Understanding*, 72(2): 143–162.

DEBEVEC, P. E., TAYLOR, C. J. and MALIK, J., 1996. Modelling and rendering architecture from photographs: a hybrid geometry- and image-based approach. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 11–20.

FRÜH, C. and ZAKHOR, A., 2001. 3D model generation for cities using aerial photographs and ground level laser scans. *Computer Vision and Pattern Recognition,* 2: 31–38.

GRUEN, A., 1985. Adaptive least square correlation: a powerful image matching technique. *South Africa Journal of Photogrammetry and Remote Sensing*, 14(3): 175–187.

GRUEN, A., 1998. TOBAGO—a semi-automatic approach for the generation of 3-D building models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(2): 108–118.

GRUEN, A. and WANG, X., 1998. CC-Modeler: a topology generator for 3-D city models. *International Archives of Photogrammetry and Remote Sensing*, 32(4): 188–196.

GRUEN, A., ZHANG, L. and WANG, X., 2003. Generation of 3D city models with linear array CCD-sensors. *Optical 3D Measurement Techniques VI* (Eds. A. Gruen and H. Kahmen), 2: 21–31.

HAALA, N. and BRENNER, C., 1998. Interpretation of urban surface models using 2D building information. *Computer Vision and Image Understanding*, 72(2): 204–214.

HARTLEY, R. and ZISSERMAN, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge. 607 pages.

HEUVEL, F. A. VAN DEN, 2000. Trends in CAD-based photogrammetric measurement. *International Archives of Photogrammetry and Remote Sensing*, 33(5/2): 852–863.

NAKAGAWA, M. and SHIBASAKI, R., 2001. Study on making city model with image and laser range data. *22nd Asian Conference on Remote Sensing*, Singapore, 2: 1108–1112.

ROTTENSTEINER, F. and BRIESE, C., 2003. Automatic generation of building models from lidar data and the integration of aerial images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W13): 174–180.

SOHN, G. and DOWMAN, I., 2003. Building extraction using lidar DEMs and Ikonos images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W13): 167–173.

STILLA, U. and JURKIEWICZ, K., 1999. Reconstruction of building models from maps and laser altimeter data. Selected papers from the International Workshop on Integrated Spatial Databases, Digital Images and GIS. *Lecture Notes on Computer Science*, 1737: 34–46.

STILLA, U., SOERGEL, U. and JAEGER, K., 2000. Generation of 3D-city models and their utilisation in image sequences. *International Archives of Photogrammetry and Remote Sensing*, 33(B2): 518–524.

WANG, X. and HANSON, A. R., 2001. Surface texture and microstructure extraction from multiple aerial images. *Computer Vision and Image Understanding*, 83(1): 1–37.

WU, J., ZHANG, Z. and ZHANG, J., 2003. City modelling by videogrammetry based on three video image sequence strips taken from a helicopter. *SPIE*, 5286: 1003–1008.

ZHANG, Y., ZHANG, Z. and ZHANG, J., 2004. Deformation visual inspection of industrial parts with image sequence. *Machine Vision and Applications*, 15(3): 115–120.

ZHANG, Z., WU, J. and ZHANG, J., 2003. A method for determining orientation parameters in image based reconstruction of architectural scenes. *Geomatics and Information Science of Wuhan University*, 28(3): 265–271 (in Chinese).

Résumé

La reconstitution tri-dimensionnelle (3D) et la figuration de la texture des bâtiments ou d'autres constructions humaines sont les éléments clés des paysages urbains 3D. On propose de traiter la réalisation de modèles 3D des bâtiments intégrant la texture, par des techniques de photogrammétrie numérique, par une démarche efficace allant de l'approché au très fin. Pour cela on a utilisé en entrée *trois séquences d'images saisies en hélicoptère avec une caméra numérique vidéo, comprenant deux vues obliques des murs des bâtiments et une vue verticale de leurs toits. On a également utilisé comme source d'informations des données lidar et une cartographie approchée 2D numérique en format vecteur servant à la navigation automobile. Pour obtenir des valeurs initiales sur les paramètres de la caméra a chaque image, on a effectué une aérotriangulation automatique (AAT) spécialement adaptée à une séquence d'images à fort recouvrement. Pour extraire avec précision tous les segments des images, on utilise une stratégie allant de l'approché au fin qui permet d'établir les correspondances entre les silhouettes des bâtiments et les détails linéaires qui leur appartiennent, tout au long des séquences d'images. On s'assure de la stabilité et de la précision de la reconstitution en se servant d'une compensation hybride par faisceaux, mêlant points et segments. On visualise une reconstitution réaliste des bâtiments avec superposition de toute la finesse de leur texture, sous forme de modèles numériques des élévations (DEM) et d'ortho-images. Les résultats des essais effectués ont montré que la solution proposée pour la réalisation des modèles urbains 3D avait un avenir prometteur pour de nombreuses applications.*

Zusammenfassung

Zentrale Bestandteile von dreidimensionalen Stadtlandschaften sind die dreidimensionale geometrische Rekonstruktion von Gebäuden und anderen künstlichen Objekten, und eine realitätsnahe Texturierung der Oberflächen. In diesem Beitrag wird ein effektiver Ansatz zur Generierung von 3D-Gebäudemodellen mit einer Texturierung auf der Basis von Techniken der Digitalen Photogrammetrie vorgestellt. Als Eingabedaten stehen drei Bildsequenzen aus einer digitalen Videokamera zur Verfügung, die in einem Helikopter montiert ist. Damit stehen zwei geneigte Ansichten von vertikalen Wänden von Gebäuden und eine Senkrechtaufnahme der Dachflächen zur Auswertung zur Verfügung. Als weitere Informationsquellen dienen Lidardaten und eine stark generalisierte zweidimensionale digitale Vektorkarte aus der Fahrzeugnavigation. Eine automatische Aerotriangulation (AAT), die für die Bearbeitung von stark überlappenden Bildsequenzen geeignet ist, wird zur Bestimmung der genäherten Kameraparameter eines jeden Bildes eingesetzt. Um genaue Linien in den Bildern zu erhalten, wird der Bezug zwischen den Umrissen eines Gebäudes und den zugehörigen Linien in den Bildsequenzen in einer Strategie von grob nach fein ermittelt. Die Stabilität und Genauigkeit der Rekonstruktion stützt sich auf eine hybride Bündelausgleichung auf der Basis von Punkten und Linien. Die rekonstruierten Gebäude werden mit einer hochauflösenden Textur versehen und zusammen mit einem Digitalen Höhenmodell (DEM) und einem Orthobild realitätsnah visualisiert. Experimentelle Ergebnisse zeigen, dass dieser Ansatz zur Generierung von 3D-Stadtmodellen eine vielversprechende Zukunft in vielen Anwendungen haben wird.

Resumen

La reconstrucción tridimensional (3D) y la texturización de edificios y de otros elementos construidos son procesos fundamentales en la visualización 3D de los paisajes urbanos. Se propone un procedimiento "grosero-a-fino" para generar un modelo tridimensional texturizado basado en técnicas de fotogrametría digital. Como información de entrada se utilizan tres secuencias de imágenes de vídeo, de las *cuales dos son las perspectivas oblicuas de las paredes de los edificios y una la perspectiva vertical de los tejados de los edificios, obtenidas con una cámara digital de vídeo montada en un helicóptero. También se utilizan datos lídar y un mapa vectorial digital plano de los usados en los sistemas de navegación de vehículos. Para obtener los valores iniciales de los parámetros de la cámara de cada imagen se realiza una triangulación aérea automática, adecuada para secuencias de imágenes con mucho solape. A continuación, y para obtener líneas exactas, se determina la correspondencia entre los contornos de los edificios y los elementos lineales de las imágenes mediante una estrategia "grosero-a-fino". Para asegurar la estabilidad y exactitud de la reconstrucción se usa un ajuste por haces híbrido punto/línea. Los edificios así reconstruidos se visualizan con realismo superponiendo texturas finas a un modelo digital de elevaciones (MDE) y a una ortofoto. Los resultados experimentales muestran que el enfoque propuesto para la generación de un modelo urbano tridimensional tiene un futuro prometedor en muchas aplicaciones.*

